



Grouped Goodness-of-Fit Tests for Binary Regression Models

by

Jana Dorteia Canary

BS (Mathematics), MS (Forest Ecology)

**Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy
(Biostatistics)**

University of Tasmania, December 2013

Supervisors

Doctor Stephen Quinn

Associate Professor Leigh Blizzard

Professor David Hosmer

Research Supervisor

Professor Ronald Barry

Dedication

To my parents Willa and Jim, my husband Mark, and my daughter Jacqueline

Declaration of Originality

This thesis contains no material which has been accepted for a degree or diploma by the University or any other institution, except by way of background information and duly acknowledged in the thesis, and to the best of my knowledge and belief no material previously published or written by another person except where due acknowledgement is made in the text of the thesis, nor does the thesis contain any material that infringes copyright.

Signed: Date:

Statement of Authorship

This thesis can be made available for loan. Copying of any part of this thesis is prohibited for two years from the date this statement was signed; after that time limited copying is permitted in accordance with the *Copyright Act 1968*.

Signed: Date:

Acknowledgements

I would like to first thank my primary supervisor, Dr. Stephen Quinn, for his support. I feel fortunate to have had Steve as a supervisor. He was always willing to take time to work with me, to give carefully thought out comments on my work, and to deal with the challenges of working with someone on the other side of the globe. I am very appreciative of his efforts in guiding me through the thesis process.

I would also like to thank my other supervisors. First, thank you to Dr. Leigh Blizzard, who made it possible for me to come to Menzies and helped me to navigate through the steps of setting up a Ph.D. candidature remotely from Alaska. He also provided useful comments on my work and provided financial support through an NHMRC grant that, along with an Australian Postgraduate Award, made my research possible. I also would like to thank Dr. Ron Barry of the University of Alaska, who was kind enough to meet with me to discuss my work while I was in Alaska, particularly during the early stages of my candidature. Lastly, I would like to thank Dr. David Hosmer, who is an emeritus professor and an expert in the field of biostatistics. He kindly offered his time and advice, and I very much appreciate his help.

I would also like to thank several fellow students and staff (in no particular order) at Menzies whose support both in person and electronically made the process so much easier: Kara, Laura, Dawn, Peta, Oliver, Kathy, Karen, Petr, Barbara, Tracey, Steve, David, and Ben. Thank you for your advice, support, encouragement, and friendship. Also my friends in Alaska and Hobart who provided childcare when I needed it, chats, and support: Petra, John, Gordon, Nancy, Jay, Donie, Peter, Margaret, Trusten, William, Leslie, and Sharon.

Also, I would like to thank the two anonymous examiners of my thesis. Their thorough review and thoughtful comments were very helpful and encouraging.

Lastly I would like to thank my husband Mark Conde and daughter Jacqueline. I cannot thank Mark enough for his financial and emotional support, and for lively discussions about mathematics. And thank you to Jacqueline, for the backrubs, hugs, and understanding when I did not have time to do fun things. I hope to make it up to you soon.

Abstract

How well a proposed regression model fits the observed outcome data is a critical question. The answer may influence model selection, and the conclusions drawn. Summary goodness-of-fit (GOF) statistics are used to assess model fit. Pearson's chi-squared GOF statistic (X^2) is used to evaluate the fit of logistic regression models, but X^2 isn't appropriate when the model contains continuous covariates. Other GOF statistics are applicable, including the Hosmer-Lemeshow (HL), Pigeon-Heise (J^2), and Tsiatis (T) statistics. All have similarities to X^2 and group data artificially.

Simulation studies assessing new GOF statistics for logistic models with continuous covariates often include HL for comparison. We know of no study that compares HL , J^2 , and T . We did so here, applying the same grouping method (deciles-of-risk) to all. Our results indicated that HL and T followed their reported distributions, but J^2 did not. Its distribution was closer to $J^2 \sim \chi^2(G-2)$, where G =groups, rather than the reported $\chi^2(G-1)$. Assuming $J^2 \sim \chi^2(G-2)$, T maintained the Type I error rate twice as often as HL and J^2 . The rates of HL and J^2 were often lower than expected when dichotomous, quadratic, or interaction terms were included. The statistics had similar power to detect departures from a true underlying model.

The logistic model is the canonical generalized linear model (GLM) for binomial outcomes. Although many GOF statistics have been developed for logistic models, there are fewer for non-canonical GLM with binomial outcomes. The properties of the logistic model make the development of GOF statistics relatively straightforward, but it can be more difficult for non-canonical GLMs.

We considered whether HL , J^2 , and T could be applied to non-canonical GLM with Bernoulli outcomes and continuous covariates. Our investigation found that HL and J^2 can be applied directly, but T cannot. We introduced an augmented version of the Tsiatis model and

generalised T , ($T_{\mathcal{G}}$). We showed that under non-canonical links, $T_{\mathcal{G}} \sim \chi^2(G)$. In a second simulation study, HL , J^2 , and $T_{\mathcal{G}}$ were used to evaluate the fit of probit, log-log, complementary log-log and log binomial models. The deciles-of-risk method was applied. Type I error rates were consistently maintained by $T_{\mathcal{G}}$, while those of HL and J^2 were often lower than expected if the model included dichotomous, quadratic, or interaction terms. Because the distributions of HL and J^2 varied, it was unclear how their degrees-of-freedom could be adjusted. The statistics had similar power to detect an incorrect model in most situations. An exception occurred when a log model was incorrectly fit to data generated from a logistic model; here $T_{\mathcal{G}}$ had more power than HL or J^2 .

Table of Contents

Grouped Goodness-of-Fit Tests	1
for Binary Regression Models	1
Declaration of Originality	3
Statement of Authorship	4
Acknowledgements	5
Abstract	6
Table of Contents	8
List of Tables	12
List of Figures	14
Chapter 1 Introduction	15
1.1 Background	15
1.2 Research Questions	20
1.3 Organization of Thesis	21
Chapter 2 Notation and Basic Concepts	22
2.1 Notation	22
2.2 Generalized Linear Models	22
2.3 Exponential Family	23
2.4 GLM for Binary Data	27
2.5 Canonical GLM for Bernoulli Outcomes (Binary Logistic Regression)	30
2.6 Non-Canonical Link Functions for Binary Outcomes	31
2.6.1 Probit	31
2.6.2 Log-log and Complementary Log-log	32
2.6.3 Log Binomial	34
2.7 Basic Concepts of Score Tests	35
Chapter 3 Literature Review	38
3.1 Goodness-of-Fit Statistics for Binary Logistic Regression Models	38
3.1.1 Deviance	38
3.1.2 Pearson's Chi-Squared	39

3.2	Goodness-of-Fit Statistics for Binary Logistic Regression Models with Continuous Covariates	40
3.2.1	Hosmer-Lemeshow Goodness-of-Fit Statistic	41
3.2.2	Tsiatis Goodness-of-Fit Score Statistic	44
3.2.3	Pigeon-Heysel Goodness-of-Fit Test Statistic	47
3.3	Other GOF Statistics for Logistic Models with Continuous Covariates	50
3.3.1	Goodness-of-Fit Statistics with Grouping Methods Based on Clustering	50
3.3.2	Smoothing Methods for Testing the Fit of Logistic Regression	51
3.3.3	Goodness-of-Fit Statistics for Logistic Models with Discrete Covariates	52
3.3.4	Score Tests for Assessing the Fit of Logistic Regression Models	54
3.4	Studies Comparing the Performance of GOF Statistics for Binary Logistic Regression Models	57
3.5	Goodness-of-Fit Statistics for Non-Canonical GLM	59
3.5.1	Statistics to Assess the Fit of Non-Canonical GLM with Discrete Covariates	59
3.5.2	Goodness-of-Fit Statistics for Assessing the Fit of Probit Models with Continuous Covariates	61
3.5.3	Assessing the Fit of Log Binomial Models.....	62
Chapter 4	Comparison of HL , J^2 , and T when Assessing the Fit of Logistic Models	63
4.1	Introduction.....	63
4.2	Algebraic Comparison	64
4.2.1	Hosmer-Lemeshow Goodness-of-fit Statistic	64
4.2.2	Pigeon-Heysel Goodness-of-fit Statistic	65
4.2.3	Tsiatis Goodness-of-Fit Statistic	65
4.3	$HL \leq J^2$	67
4.4	J^2 Can Be Much Larger Than HL	69
4.5	Simulation Study Comparing HL , J^2 and T	70
4.5.1	Simulation Methods	71
4.5.1.1	General Simulation Methods	71
4.5.1.2	Methods to Investigate the Null Distributions of HL , J^2 , and T	72
4.5.1.3	Methods for Comparing of the Null Empirical Rejection Percentages of HL , J^2 , and T	75

4.5.1.4	Methods for Comparing the Power of HL , J^2 , and T	76
4.5.2	Simulation Results	77
4.5.2.1	Distribution of J^2	77
4.5.2.2	Empirical Rejection Percentage Under the Null Hypothesis	80
4.5.2.3	Power - Rejection Percentage Under the Alternative Hypothesis.....	83
4.6	Examples.....	86
4.7	Discussion	88
Chapter 5 Proposed Goodness-of-Fit Statistic for Binary GLM with Non-Canonical Links...		91
5.1	Expanded Tsiatis model	91
5.2	Generalized Tsiatis GOF Statistic.....	93
5.3	Forms of $T_{\mathcal{G}}$ Under Several Common Link Functions	96
5.4	Distribution and Degrees of Freedom of $T_{\mathcal{G}}$	97
5.5	Grouping Method.....	101
5.6	Examples of Alternative Tsiatis and Generalized Tsiatis Models	102
5.7	HL and J^2 for Binary GLM with Non-Canonical Links.....	103
5.8	Simulation Study Comparing HL , J^2 , and $T_{\mathcal{G}}$ Under Non-Canonical Links.....	104
5.8.1	Simulation Methods	104
5.8.1.1	General Simulation Methods	104
5.8.1.2	Investigation of Null Distribution of HL , J^2 , and $T_{\mathcal{G}}$	105
5.8.1.3	Empirical Rejection Percentage Under the Null Hypothesis	106
5.8.1.4	Power	113
5.8.2	Simulation Results	114
5.8.2.1	Distribution of HL , J^2 , and $T_{\mathcal{G}}$ Under Non-Canonical Link Functions.....	114
5.8.2.2	Empirical Rejection Percentage Under the Null Hypothesis	118
5.8.2.2.a	Probit.....	127
5.8.2.2.b	Log-log.....	127
5.8.2.2.c	Complementary Log-log	128
5.8.2.2.d	Log	129
5.8.2.3	Power	129

5.8.2.3.a	Probit.....	134
5.8.2.3.b	Log-log.....	134
5.8.2.3.c	Complementary Log-log	135
5.8.2.3.d	Log	136
5.9	Examples.....	136
5.10	Discussion	139
Chapter 6	Overall Discussion	142
6.1	Overview of the Chapter	142
6.2	Broad View of Research	143
6.3	Need For This Research.....	144
6.4	Contribution and Significance of This Research	145
6.5	Limitations of This Research	147
6.6	Future Research	148
Appendix A	Derivation of Terms for the Calculation of $T_{\mathcal{G}}$	150
A1	Canonical Logit Link	150
A2	Non-Canonical Links	150
A2.1	Probit Link (T_{Pr}).....	150
A2.2	Log-log Link (TLL)	152
A2.3	Complementary Log-log Link (T_{CII})	154
A2.4	Log Link (T_{LB})	156
Bibliography	158